International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 3, Issue 4, April 2016

Privacy Preserving Association Rules Mining In Horizontally Distributed Databases

[1] Nikhil Puri [2]Swapnil Patke, [3]Vishwajeet Godage, [4] Yogesh Sherkar, [5]Asst.prof.Sunil Kale
[1][2][3][4][5]Department of Computer Engineering, Rajarshi Shahu College of Engineering,
[1] nik6492@gmail.com [2]Patke.swapnil123@gmail.com [3]visha13messi@gmail.com
[4]yogeshsherkar@gmail.com[5]kalesunild@gmail.com

Abstract: **- This paper presents secure mining of association rules in horizontally distributed databases by proposing a protocol. This protocol is made up of the Fast Distributed Mining (FDM) algorithm which is consist of unsecured version of the Apriori algorithm. Our protocol made up of two novel secure multi-party algorithms — one that calculates the unions of private subsets that each of the interacting players hold, and another that tests the inclusion of an element held by one player in a subset held by another. The main merit of our protocol is to provide privacy with respect to the protocol. It is also useful in terms of simplicity and is also significantly more efficient in terms of communication rounds, communication cost and computational cost.**

***Keywords*— secure data mining, privacy preserving distributed computation, frequent patterns, association rules, Apriori and FDM algorithm, multi-party algorithm.**

I. INTRODUCTION

We are going derive secure mining of association rules in horizontally distributed databases by proposing a protocol. Proposed protocol is based on the Fast Distributed Mining (FDM) algorithm which is an unsecured distributed version of the Apriori algorithm. Successful organizations view such databases as important pieces of the marketing infrastructure. We provide new, simple, and practical constructions of message authentication schemes based on a cryptographic hash function[3].They require an institution of information-driven marketing processes, managed by database technology, that enable marketers to build and implement customized marketing schemes and strategies [1]. The main parts of our protocol are two novel secure multi-party algorithms — one that computes the unions of private subsets that all of the interacting players hold, and another that tests the inclusion of an element possessed by one player in a subset held by another. It is easy to understand why people want to MAC with cryptographic hash functions: the popular hash functions are efficient than block ciphers; these software implementations are readily and freely available; and the functions are not meant to the export prohibition rules of the USA and other countries. The more difficult question is how best to do it [2,3]. In distributed mining, synchronization happens to be implicit in message passing, so the goal becomes communication optimization [10]. Our protocol offers enhanced privacy with respect to the protocol. In addition, it is simpler and significantly has more efficiency in terms of communication rounds, communication cost and computational cost.

II. RELATED WORK

Many protocol for protected mining of association rules in horizontally scattered database are recently proposed. The current leading set of rules is of Kantarcioglu and Clifton. Our procedure, like theirs, is based on the Fast Distributed Mining algorithm of Cheung et al, which is an unsecured spread version of the Apriori algorithm. The important parts in our procedure are two novel secure multi-party algorithms one that computes the union of private subsets that all of the interacting group of actors hold, and another that tests the inclusion of an element held by one actor in a subset held by another [4]. This protocol offers improved separation with respect to the protocol and it is simpler and extensively more efficient in terms of announcement rounds, announcement cost and computational cost [4,9]. The aim is to discover all association rules with support at least s and confidence at least c, for some given minimum support size s and confidence level c, that hold by the unified database, while reducing the information disclosed about the confidential databases held by those players [7,8]. The information that we would like to secure in this context is not only individual transactions in the given different databases, but also more global information such as what association rules are locally supported in each of those databases [6,7].

III. PROPOSED SCHEME

The protocol that we propose here processes a parameterized group of functions, which we call edge capacities, in which the two compelling cases compare to the issues of figuring the union and crossing point of private subsets. Those are truth be told universally useful conventions that can be utilized as a part of different connections also. Another issue of secure multiparty calculation that we explain here as a component of our discourse is the set incorporation issue; to be specific, the issue where Alice holds a private subset of some ground set, and Bob holds a component in the ground set, and they wish to figure out if Bob's component is inside of Alice's subset, without uncovering to both of them data about the other party's information past the above depicted consideration.

IV. SYSTEM ARCHITECTURE

System architecture gives the flow of data inside the system. It has various phases as shown in fig.1. First phase initialization, in which the player start their role by holding some value in it. And then it will help to find out the next item. Second phase is generating candidate set, in which we are finding the key which appears repeatedly or which is common for both sites and players. Third phase is local pruning, in which we are trying to discard the unwanted result or extra data which will in turn help in mining the data. Fourth phase is unifying Candidate item sets, as word indicates it is based on the union of data sets of participating players. Fifth phase is local support computation, in which we are computing the local support that how much the all participating players can support. Sixth phase is Broadcasting of the mining result in which we display the result by combining the all results that we got from all participating player and then displaying it.

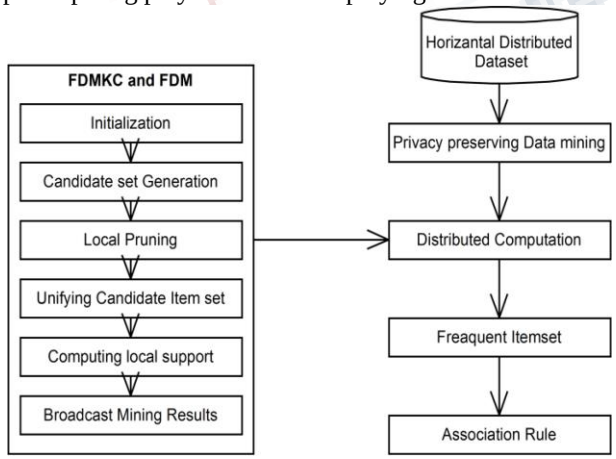

Fig 1: System architecture

V. MATHEMATICAL MODEL

Input:-

Let S is the Whole System Consist of

S= {U, FS, IS, ICF, DROP, POP, IG, SU, RF, O}

WHERE,

IS = INSTANCE SELECTION.

FS = FEATURE SELECTION.

ICF = ITERATIVE CASE FILTER.

DROP = Decremental Reduction Optimization Procedure.
POP = Patterns by Ordered Projections.
IG = Information Gain
SU = Symmetrical Uncertainty
RF = Relief-F Attribute selection

Procedure:-

Step1: Admin send the data reduction request.
Step2: Instance Selection operation performed by system.
Step3: Feature selection operation performed by system.
Step4: Classification.
Step5: Creation of new bug data-set.
Step6: Admin will assign that bug to respective Developer.
Step7: Developer will remove that bug.
Instance and Feature Selection:-
FS → IS = Bug data reduction.
Which first applies FS and then IS on the other hand, IS → FS denotes first applying IS and then FS In Algorithm 1, we briefly present how to reduce the bug data based on FS → IS.?
Given a bug data set, the output of bug data reduction is a new and reduced data set.
Two algorithms FS and IS are applied sequentially
Note that in Step2, some of bug reports may be blank during feature Selection
In our work, FS → IS and IS → FS are viewed as two orders of bug data reduction.
To avoid the bias from a single algorithm, we examine results of four typical algorithms of instance selection and feature selection, respectively.
1. Iterative Case Filter (ICF)
2. Learning Vectors Quantization (LVQ)
3. Decremental Reduction Optimization Procedure (DROP)
4. Patterns by Ordered Projections (POP)

Output: Getting the appropriate Decision from the above technique **.**

IV. CONCLUSION

In our dissertation we aim to develop privacy preserving protocol. We extend secure mining protocol used for distributed databases. We believe that this will improve security and privacy of mining operations in distributed database. We proposed a system that will help for secure mining of data in horizontally distributed databases. In our base paper we implement a secure protocol for mining of association rules in horizontally distributed database. We

aim to extend this work by developing privacy preserving protocol using cryptography. This is because, in case of distributed databases when databases don't trust each other's association rules mining is difficult as honest nodes may lose privacy.

REFERENCES

[1] R.Agrawal and R.Srikant, "Fast algorithms for mining association rules in large databases," In VLDB, pages 487–499, 1994.

[2] D.Beaver, S. Micali, and P.Rogaway, "The round complexity of secure protocols," In STOC, pages 503–513, 1990.

[3] M.Bellare, R. Canetti, and H. Krawczyk, "Keying hash functions for message authentication," In Crypto, pages 1–15, 1996.

[4] A. Ben-David, N. Nisan, and B. Pinkas, "FairplayMP - A system for secure multi-party computation" In CCS, pages 257–266, 2008.

[5] J.Brickell and V.Shmatikov, "Privacy-preserving graph algorithms in the semi-honest model," In ASIACRYPT, pages 236–252, 2005.

[6] Rajkumar.S, V.Elavaras, "A fast distributed algorithm for mining association rules," In PDIS, pages 31–42, 1996.

[7] Tamir Tassa, D.W.L Cheung, V.T.Y. Ng, A.W.C. Fu, and Y. Fu., "Efficient mining of association rules in distributed databases," IEEE Trans. Knowl. Data Eng., 8(6):911–922, 1996.

[8] S.Pavithra, P.Prasanna, "A novel secure multiparty algorithms in horizontally distributed database for fast distributed database," 24:106–110, 1978.

[9] Bollu Jyothi, K.Venkateswara Rao, "Secure Mining of Association Rules in Horizontally Distributed Databases" (Protecting Sensitive Labels in Social Network Data Anonymization) pages 639–644, 2002.

[10] N.Kowsalya, M. Saraswath, "Privacy Preserving and Secure Mining of Association Rules in Distributed Data Base," pages 139–147, 2005.